

**PROTOCOL TITLE : Randomized, double blind, placebo controlled, multicenter pilot study on the effects of empagliflozin on clinical outcomes in patients with acute decompensated heart failure (EMPA-RESPONSE-AHF)**

**September 2019**

NCT03200860

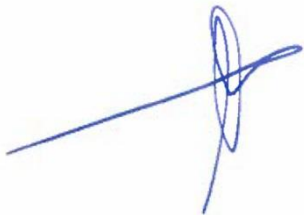
# EMPA-RESPONSE-AHF

---

## Statistical Analysis Plan

Version 1.0

September 2019



---

Prof. dr. A.A. Voors



---

Prof. dr. H.J. Lambers Heerspink

## 1. Population for analysis

The following analysis population will be defined for statistical analysis:

- a) The Full Analysis Set (FAS) population consists of all randomized patients. Following the Intention To Treat (ITT) principle, all randomized patients will be analyzed, according to the treatment they were assigned to at randomization
- b) Per Protocol (PP) population consists of a subset of FAS, will consist of all randomized patients without major protocol deviations.
- c) Safety Population will consist of the FAS population and includes patients that received at least one dose of the study drug and had one safety assessment thereafter

The co-primary efficacy variables will be analyzed using the FAS population. All primary analysis will be repeated for the per protocol population. All secondary analyses will be analyzed using the FAS population

## 2. Demographics

Demographics of the population will be summarized in a baseline characteristics table stratified for treatment group. Continuous variables will be summarized using mean, median, standard deviation, 25<sup>th</sup> and 75<sup>th</sup>, depending on distribution. Categorical variables will be summarized using frequency and percentage. Characteristics include (but not limited to): Age, sex, race, height, weight, body mass index (BMI), left ventricular ejection fraction (LVEF), systolic and diastolic blood pressure, heart rate, NYHA class, history of: heart failure, myocardial infarction, diabetes mellitus, atrial fibrillation, hypertension, cerebrovascular disease, medical therapy including medication and devices, and laboratory measurements at baseline, including hemoglobin, hematocrit, sodium, potassium, serum creatinine/eGFR, NTproBNP.

Treatment groups will be compared using Chi-square or Fisher's exact test for categorical variables and t-test or Wilcoxon rank-sum test for continuous normal and non-normally distributed variables. P-values will be provided for descriptive purposes. For exploratory analyses, if an imbalance in any of the variables between groups exist, adjustment for these variables in the model may be considered.

## 3. Analysis of the Primary Endpoints

The primary endpoints of this study are:

- Change in Dyspnea VAS from Baseline to Day 4
- Diuretic response (defined as  $\Delta \text{ weight kg} / [(\text{total i.v. dose})/40\text{mg}] + [(\text{total oral dose})/80\text{mg}]$  furosemide (or equivalent loop diuretic dose))
- Length of initial hospital stay
- Change in NTproBNP from baseline to Day 4

As this is an exploratory study with limited power, all 4 individual endpoints will be tested separately, with no formal correction for multiplicity. As sensitivity analysis, if at least 2 out of 4 primary endpoints show significant difference in the same direction (favoring either investigational drug or placebo) a Bonferroni correction will be applied. The significance level for the primary outcomes will

be a conservative two-tailed  $P < 0.2$ . With Bonferroni correction sensitivity analysis, the significance level will be  $P < 0.05$ .

The difference between treatments in the change in Dyspnea VAS score from Baseline to Day 4 will be assessed by comparing the area under the curve (AUC) of VAS score over time by student's t test. To do so, individual changes in VAS score will be visualized (virtually) as a curve where the X-axis shows study day baseline to day 4, and y-axis shows VAS score. Using this approach, area under the curves for each study day (trapezoids) can be calculated, and added together, resulting in an overall VAS AUC score (mmxh) that can be compared across treatment groups. This method follows methods described by the DOSE and RELAX-AHF trials.

The difference in Diuretic response through day 4 will be assessed using either students t test if normally distributed or Wilcoxon rank-sum test if non-normally distributed. Diuretic response will be calculated as the change in weight from baseline to day 4 divided by the total dose of administered furosemide (total intravenous dose /40 mg + total oral dose/80 mg). Bumetanide will be recalculated to furosemide equivalents by the following equation: 1 mg of Bumetanide = 40 mg Furosemide.

Difference in Length of stay will be assessed using either students t test if normally distributed or Wilcoxon rank-sum test if non-normally distributed.

The difference in percentage change in NTproBNP from baseline to day 4 will be assessed using students t test if normally distributed or Wilcoxon rank-sum test if non-normally distributed. Secondly, using all NTproBNP levels at all timepoints through day 4, using a mixed effect repeated measures model (covariates including randomized treatment, study day, interaction between randomized treatment and study day) the interaction between change in NTproBNP over time and randomized treatment will be evaluated. NTproBNP will be log transformed if non normally distributed for this analysis.

As an exploratory analysis, and for graphical presentation, individual responses to the above mentioned endpoints will be standardized. This will be done by dividing the difference of the overall mean (or log transformed mean if non normally distributed) of each endpoint by the overall standard deviation (SD) of that variable. The treatment effect can then be measured by the mean difference of standardized scores, which can be visually presented on a forest plot by mean  $\pm$  80% CI (given the  $P < 0.2$ ).

Then, all four standardized scores for each individual endpoints will be averaged, and mean treatment difference and associated 95% CI for this overall treatment effect visually presented. Statistical analysis for the treatment difference will then be carried out by students t-test.

The primary analysis will be repeated for the per-protocol population as supportive.

#### **4. Analysis of the Secondary Endpoints**

The secondary endpoints of the study are:

- Death through day 30
- Heart Failure Re-admission through day 30
- Death or Heart Failure re-admission through day 30
- Change in cardiorenal biomarkers from baseline to day 4 or day 30

The difference in the incidence of the mortality/rehospitalization endpoints between both randomized treatment groups will be assessed using logistic regression. Crude event rates, percentages per treatment group and odds ratio's with 95% confidence intervals will be presented. Changes in cardiorenal biomarkers and the difference in change between treatment groups, including plasma or urinary biomarkers will be assessed using t-test for normal distributed variables and Wilcoxon rank sum test if non-normally distributed for change between two timepoints. For change over time at multiple timepoints a mixed effect repeated measures model (covariates including randomized treatment, study day, interaction between randomized treatment and study day) will be carried out. Alternatively, an ANCOVA model using similar approach may be used.

## 5. Analysis of the Safety Endpoints

Analysis of safety endpoints will be conducted in the Safety Population. Adverse events include:

- Adverse events (general)
- Adverse events that lead to treatment discontinuation
- Serious Adverse Events (which could include a secondary endpoint)
- Adverse Events of Special Interest (AESI), including:
  - o Hepatic Injury
  - o Worsening Renal Function
  - o Metabolic Acidosis, Ketoacidosis and Diabetic Ketoacidosis

Overall adverse event rates in the treatment groups will be presented as counts (percentages) and will be compared using Fisher's exact test.

## 6. Subgroup analysis

As this is a small randomized controlled trial, subgroup analysis are supportive and exploratory in nature. Subgroup analyses will be performed in the FAS population. The primary outcome variables, it's individual components and the mean treatment score, will be evaluated using the primary analysis and adding the specific subgroup as a factor and treatment x subgroup as interaction term.

The following subgroups will be analyzed, no adjustment for multiple comparisons will be carried out:

- Age groups
- Gender
- De novo versus acute decompensated (known) heart failure
- Heart failure with preserved versus reduced ejection fraction
- eGFR groups (above/below median baseline values)
- History of Diabetes Mellitus (yes/no)
- Blood pressure groups
- NTproBNP at admission (above/below median levels)

## 7. Sample Size and Power calculation

Our sample size calculation is based on capturing the primary outcome measures in the placebo group with a degree of certainty. With 40 patients in the control group, a given mean continuous response can be estimated within  $\pm 0.2$  Standard deviations (SD) with 80% confidence. We estimate the following mean responses for the primary outcome measures (in the placebo group):

1. Change in dyspnea VAS score:  $1756 \pm 2353$  mmxh,
2. Diuretic response (of  $0.56 \pm 0.78$  kg/40 mg Furosemide (or equivalent))
3. Length of stay 9.6 to 10.5 days ( $\pm 9.1$ )
4. Percentage change in NT-proBNP 24 ( $-1.0 - 88.7$ )% (SD 67% assuming normal distribution) (all at 96 hours (day 4))

For the Secondary outcomes, including death and/or HF re-admission at 60 days, an event rate of 10% can be estimated within  $\pm 6.1\%$  and a rate of 15% within  $\pm 7.2\%$  with 80% confidence.

### Treatment effect:

As this is an exploratory analysis the overall power to establish small treatment effects is limited. Forty patients per group provides approximately 80% power to detect standardized mean treatment differences of approximately 0.48 SDs at the two-sided 20% significance level, which gives around the following differences:

1. Dyspnea VAS score of 1129 mmxh
2. Diuretic response of 0.37 kg/40 mg Furosemide
3. Length of initial hospital stay of 4 days
4. Percentage change in NT-proBNP of 32%

with the above SDs the control group (and corresponding mean response).